Clustering and Likelihood Based Analysis for Muon Radiography LA-UR-05-2659

by
Tom Asaki

April 14, 2005

1 Introduction

This report outlines the use of clustering and likelihood algorithms in analyzing muon radiography data for the selection of likely locations of dense high-Z material (DHZ) in a cargo container. The distance function is physically motivated. The likely locations are intended primarily as an input to existing support vector machine (SVM) classifiers developed independently from this analysis. It may prove possible, given a good density function, to also independently locate SNM in some situations. The report outlines my thought process and ends with a results section from Matlab analysis on standard synthetic data sets.

2 Define Relevant Data

To a fair approximation, muons of a given energy E passing through a thickness L of a material characterized by radiation length L_0 inherit a nearly Gaussian angular distribution of width θ_0 given by

$$\theta_0 = \left(\frac{13.6MeV}{E}\right)\sqrt{\frac{L}{L_0}}\left[1 + 0.038ln\left(\frac{L}{L_0}\right)\right]. \tag{1}$$

From Eq. 1, I define the scattering significance value S that is the energy normalized scattering width:

$$S \equiv \frac{\theta E}{13.6MeV} = \sqrt{\frac{L}{L_0}}\left[1 + 0.038ln\left(\frac{L}{L_0}\right)\right] \tag{2}$$

for a given muon of energy E scattered with a total angle θ. Similarly, define the unit significance S^* to be that associated with scattering angle θ_0.

Thus, materials that tend to provide more large-significance muons are those that are thicker and or have smaller radiation lengths. Some representative radiation lengths (given in cm) are 36(water), 1.76(Fe), 0.56(Pb), 0.353(Pu), and 0.307(U). Hopefully, it will be possible to detect the presence of DHZ by their tendency to produce high-significance muons.

Consider the following representative scenario for muon radiography. Let a shipping container (2.4m wide by 6.0m long by 2.4m high) hold 110 spherical half-density iron balls of radius 20cm. They are distributed in two layers of 55 each. This cargo represents 14 metric tons (the approximate weight limit) and covers about half of the container bed. Since the spheres are half density, the approximate radiation length is 3.5cm. Also included are one sphere of U of radius 6.33cm and one sphere of Pu of radius 4.22cm. This scenario is similar to a combined '1c' and '2c' test case of Katz and Borozdin [1]. One can calculate the number of muons N exceeding a given significance value associated with the three cargo materials in a given amount of time t and the density of detected muon scattering locations D (events per cubic meter) assuming perfect single-scatter ray track knowledge. These calculations are intended to be illustrative and possibly even representative of simple muon tracking reconstruction capability. For a one minute experiment (95000 detectable muons) I calculate the following table.

	N			D		
(S/S^*)	Fe	U	Pu	Fe	U	Pu
0	47981	86	38	25468	80569	120782
1	35352	74	31	18765	69731	99134
2	24060	63	25	12771	59200	78564
3	15032	52	19	7979	49254	59994
4	8579	43	14	4554	40127	44063
5	4455	34	10	2365	31988	31078
6	2098	27	7	1114	24933	21021
7	894	20	4	475	18992	13620
8	344	15	3	183	14130	8445
9	119	11	2	63	10263	5006
10	37	8	1	20	7274	2836
11	10	5	0	6	5029	1534
12	3	4	0	1	3391	792
13	1	2	0	0	2229	390
14	0	2	0	0	1428	183

While a large amount of muons are affected by the iron, the angular distribution decays rapidly relative to U or Pu. In order for a clustering approach to be effective, some nonnegligible number of muons must be associated with DHZ while at the same time not be overwhelmed by the number of muons associated with the iron. The expectation is that $N_{DHZ} < N_{Fe}$ but that $N_{DHZ} \ll N_{Fe}$ should be avoided when possible. The detection of a cluster can be made with as few as two or three muons. An accurate estimate of a cluster centroid depends upon many factors but certainly requires more than just a few muons, say $N_{DHZ} \geq N^* \sim 10$. Also, clustering techniques and likely-location estimation will perform better when $D_{DHZ} \gg D_{Fe}$. Finally, the real-time computational requirement coupled with the computational efficiency of clustering imposes a limit $N_{Fe} + N_{DHZ} \leq N_{comp} \sim 1000$. These clustering criteria are easily met for the U case using a significance cutoff of around nine. The Pu case seems intractable for a time of one minute. Even an experiment time of 10 minutes (data not shown, but can be inferred from the table) still seems difficult.

The cases thus far described represent somewhat challenging scenarios. Halving

the amount of iron (more representative?) would provide a significant improvement in clustering ability, especially in the difficult Pu case. The actual cutoff value could be estimated in terms of the size of the SNM to be detected, the experiment time and the cargo weight.

Now that the muons of interest have been identified (the subset whose significance value exceeds some cutoff) the data property for clustering must be identified. In this case we consider the most likely location of a single-scattering event. This location p is the point of closest approach (PoCA) of the experimentally measured incoming and outgoing muon paths illustrated in Fig. 1. The incoming muon path is determined by detector locations c and d. The outgoing muon path is determined by detector locations g and h. The shortest line segment connecting these to lines is that connecting points a and b. The center of this line segment is the PoCA location p. For all sufficiently significant muons the PoCA locations are the data set on which clustering will be performed.

The nonideal nature of the PoCA due to small angle scattering uncertainty and the single-scatter assumption is addressed in the definition of the distance function in a later section.

One might be concerned that if a small subset of the total number of muons is used then information is being lost unnecessarily. Two things need to be remembered. First, the entire collection of muons is used to define the subset on which clustering should be performed. Second, the likely locations of DHZ given by a clustering are only input locations to a DHZ analysis which does consider all muons relevant to that particular container location vicinity.

3 Formulate a Distance Function

A good distance function will respect our directional knowledge in the uncertainty in the PoCA location and the characteristic size of objects for detection. Generally we know that uncertainty is large in directions along ray paths and relatively small in orthogonal directions. I begin with a single-muon motivated coordinate system with orthonormal basis $\{\vec{e_{ji}}\} = \{\vec{e_1}, \vec{e_2}, \vec{e_3}\}$. The first index j is a muon index and the second index i is a coordinate direction index. The first basis vector is

$$\vec{e_{j1}} \equiv \frac{\vec{b} - \vec{a}}{\left|\vec{b} - \vec{a}\right|} \tag{3}$$

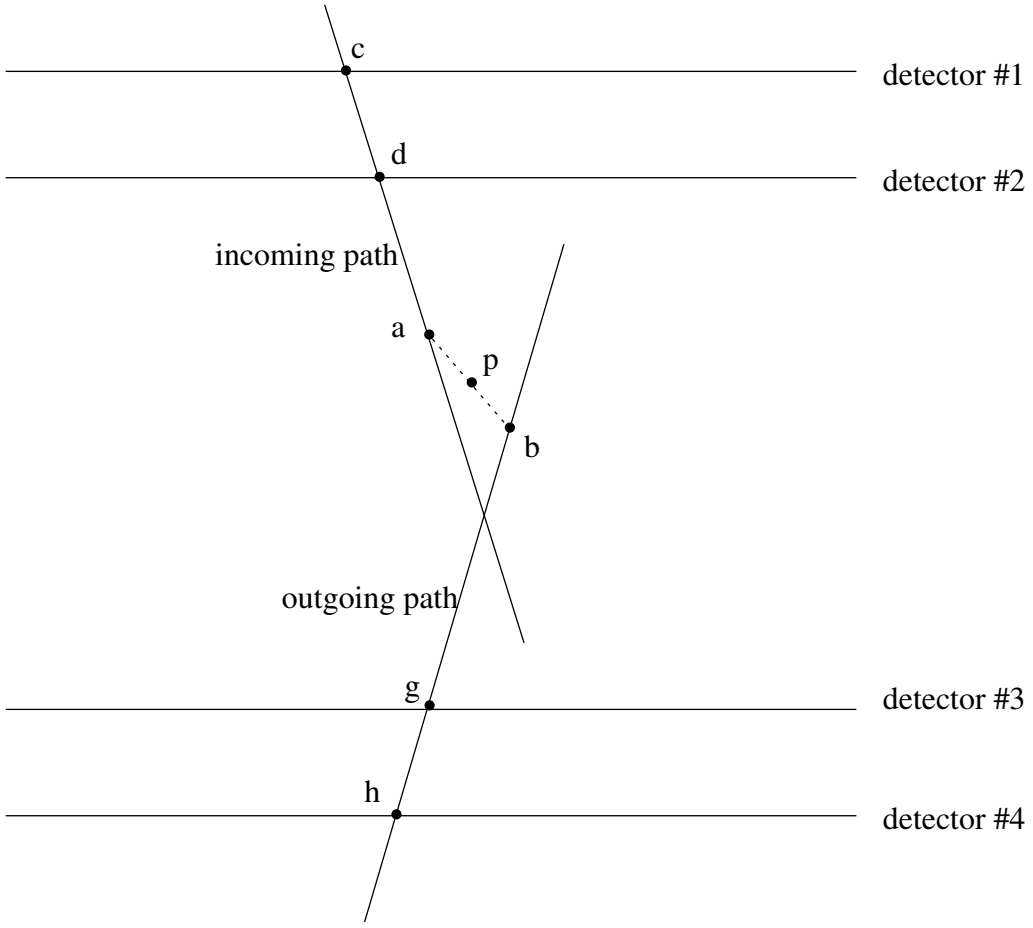

Figure 1: Schematic of the point of closest approach (PoCA) idea. All points with roman labels are 3D locations. The entering muon path is defined by c and d. The exiting muon path is defined by g and h. The PoCA location p is given by the center of the shortest line segment (endpoints a and b) connecting these two paths. The orthonormal vectors $\vec{e}_i$ are the principal directions associated with the muon path.

which is perpendicular to both rays. It is the direction from p along the line segment $\bar{ab}$. It is the direction vector that identifies the quickest possible path away from the ray paths. I have adopted vector notation to identify point locations: $\vec{a}$ is the vector locating point a relative to an arbitrary fixed origin.

Next, formulate the perpendicular basis vector that identifies the angular deflection from incoming to outgoing raypath. Let

$$\vec{r_{in}} \equiv \frac{\vec{c} - \vec{a}}{|\vec{c} - \vec{a}|} \tag{4}$$

and

$$\vec{r_{out}} \equiv \frac{\vec{g} - \vec{b}}{\left|\vec{g} - \vec{b}\right|}. \tag{5}$$

Then,

$$\vec{e_{j2}} \equiv \frac{\vec{r_{in}} + \vec{r_{out}}}{|\vec{r_{in}} - \vec{r_{out}}|}. \tag{6}$$

And finally, the third basis vector is the direction most closely directed along the two ray paths:

$$\vec{e_{j3}} \equiv \vec{e_{j1}} \times \vec{e_{j2}}. \tag{7}$$

Now a distance function can be defined relative to a single muon that can incorporate physical knowledge of the PoCA uncertainties. Let the distance from an arbitrary point s to the j^{th} PoCA location p_j be

$$d_{sj} \equiv \sqrt{\sum_{i=1}^{3} \left(\alpha_{ji} \left(\vec{s} - \vec{p_j}\right) \cdot \vec{e_{ji}}\right)^2}. \tag{8}$$

The distance is weighted in the three local coordinate directions by paramters α_{ji}. Since the basis vectors and weights are dependent upon the specific muon, the distance function is not symmetric between two PoCA location p_j and p_k. That is, $d_{jk} \neq d_{kj}$ in general. A symmetric distance metric is formed by choosing

$$d_{jk} = d_{kj} \equiv \sqrt{\sum_{i=1}^{3} \left(\alpha_{ji} \left(\vec{p_k} - \vec{p_j}\right) \cdot \vec{e_{ji}}\right)^2} + \sqrt{\sum_{i=1}^{3} \left(\alpha_{ki} \left(\vec{p_j} - \vec{p_k}\right) \cdot \vec{e_{ki}}\right)^2}. \tag{9}$$

Of course, when all $\alpha = 1$, the distance metric is proportional to the Euclidean distance. Further discussion on the choice of the α_{ji} is addressed in the section on centroid and density calculations.

A distance function is also needed to determine a centroid location s^* relative to a cluster of n PoCA locations p_k. I propose that s^* is the point providing the solution to the following minimization problem.

$$\tilde{r}_t = \min_s \frac{1}{n} \sum_{k=1}^{n} \sqrt{\sum_{i=1}^{3} \left(\alpha_{ki} \left(\vec{s} - \vec{p_k}\right) \cdot \vec{e_{ki}}\right)^2}. \tag{10}$$

Now, $\tilde{r}_t$ is identified as a characteristic length scale of the cluster t. It can be used to define a cluster density

$$\rho_t \equiv \frac{n}{\tilde{r}_t^3}. \tag{11}$$

I now turn to the question of selecting the weights α_{ji}. The physicists suggest that $\alpha_{j1} = \alpha_{j2} = \alpha_{j3}/\sqrt{3}\delta\theta$, where $\delta\theta = \arccos\left(-\vec{r_{in}} \cdot \vec{r_{out}}\right)$ is the angle between incoming and outgoing rays. Furthermore, I expect that $\alpha_{j1} \propto \left|\vec{a} - \vec{b}\right|^{-1}$ and that $\alpha_{j2} \propto \delta\theta^{-1}$.

$$\alpha_{j1} = \alpha_{j2} = \frac{1}{\left|\vec{a} - \vec{b}\right| \cdot \delta\theta}, \tag{12}$$

$$\alpha_{j3} = \frac{\sqrt{3}}{\left|\vec{a} - \vec{b}\right|}. \tag{13}$$

While α_{j3} can tend to be relatively small for typical deflection angles of several milliradians, these equations certainly reflect the idea that uncertainties are greater along ray paths.

Finally, since objects have finite length scales, it seems unwise to over-penalize any distance differences within this length scale. For some applications, I propose the use of a modified distance funciton X_{sj} given a characteristic object size a:

$$X_{sj} = d_{sj} \left[1 - exp \left(- \left(d_{sj}/a \right)^p \right) \right]. \tag{14}$$

This modified distance provides a smooth transition between distances large compared to a, where $X_{sj} \approx d_{sj}$, and an effective small distance region of radius $r \approx a$. Figure 2 shows this modified distance function for a few values of the parameter p. The choice of p will depend upon the particular application. At present I have not utilized such a distance function due to the complexity of use in potential function applications (calculating derivatives). Instead I make use of modified potentials which I present in a later section.

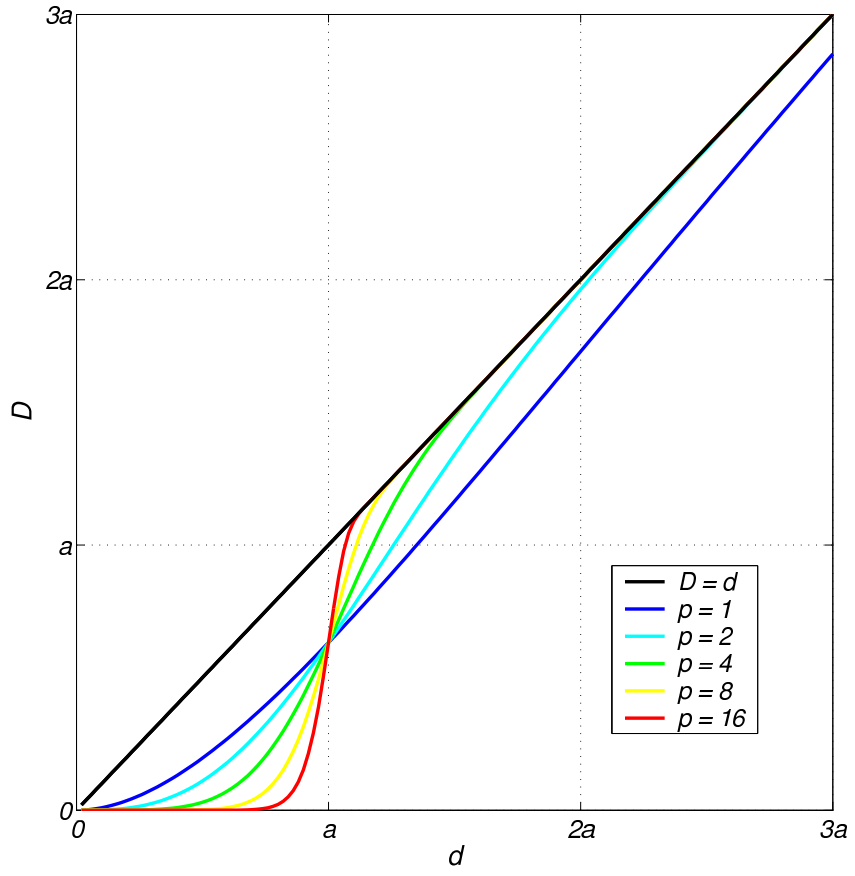

Figure 2: Modified distance function plots (see Eq. 14) for different values of p.

4 Data Clustering Approach

For this work, the primary interest is in identifying or locating compact objects of size characteristic of the PoCA location uncertainty. Objects are expected to be non-extended, solid, and of some minimum characteristic size. However, it may be important to identify embedded objects. For example, a small amount of DHZ surrounded by a relatively large amount of other material, such as iron. Thus, any data clustering algorithm should be sensitive to these geometries.

Relevant example objects include spheres, hemispheres, cylinders, rectangular blocks, and irregular 3D compact solids. Object examples falling outside our clustering scheme include extended rods, distributed small objects, powders, etc.

Because of the simplicity of the objects sought, it makes sense to use a K-means based clustering method. I am working with a dynamic membership scheme that iteratively tests the largest cluster for possible division. This process continues until stability is reached. An outer computational loop governs the selection of a division parameter γ and exits according to some stopping criteria. An additional loop allows for periodic data reclustering with freedom for data points to move between clusters if favorable. A flow chart is shown in Fig. 2. Details of the algorithm are given in the working Matlab code available from the author. The clustering algorithm uses the basic distance function, Eq. 8, finds cluster centroids according to Eq. 10, and computes cluster densities using Eq. 11.

Representative clustering results applied to six cargo container scenarios are shown in Figures 4-9. I use the basic distance function given by Equation 8 and the algorithm given in Figure 3 applied to the 100 most significant muons from a one minute simulation. Perfect muon energy knowledge is assumed.

Each of the six figures shows two views of clustering results. Subfigures (a) show the PoCA locations of the most significant muons (black dots), the cluster centroids (bold blue dots), and, if it exists, a surface rendering of the DHZ. Subfigures (b) are either close-in views of the region containing DHZ or a different 3D view of the container depending upon the scenario. The six scenarios are given by the Katz-Borozdin notation 1a 1c 2a 2c 6a 6c, respectively. Numbers indicate centrally located DHZ: 1 = 6.33cm radius sphere of uranium; 2 = 4.22cm radius sphere of plutonium; 6 = no DHZ. Letters indicate non-DHZ background material: a = no material; c = 110 half-density iron spheres of radius 20 cm placed in two layers in the lower part of the container; and e = 9 cm thick Pb slabs on top and bottom of container.

I also performed clustering calculations on 100 realizations of each of the scenarios 1a, 1c, and 1e. For each I retained up to 10 likely locations returned by the clustering algorithm. From this list I selected, for each realization, the location closest to the actual location of the DHZ. Figure 10 shows histogram plots of the number of cluster locations within a given distance of the correct location (in units of DHZ radius). The results are encouraging, but fall short of the high accuracy needed for the application. K-means based clustering results can vary significantly with cluster division and stopping criteria so that centroid locations vary by distances comparable to cluster sizes.

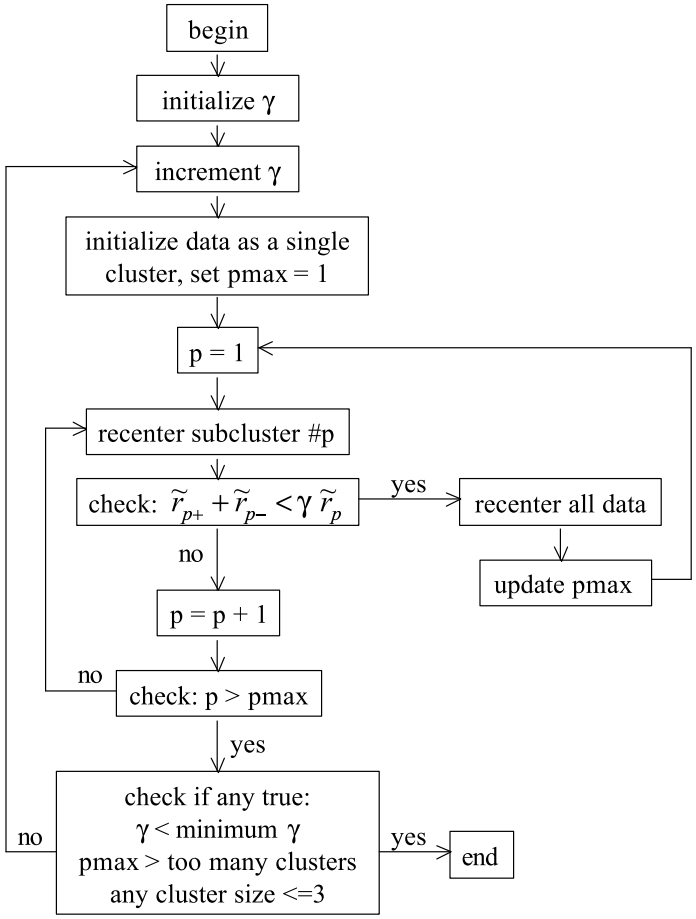

Figure 3: Flow chart of the clustering method.

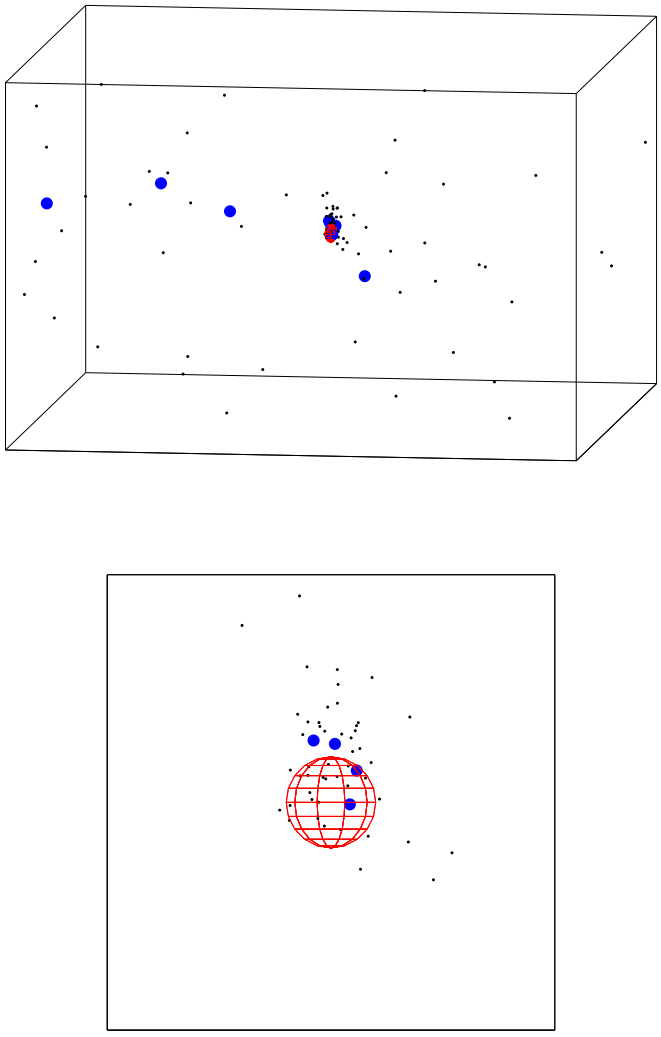

Figure 4: Scenario 1a.

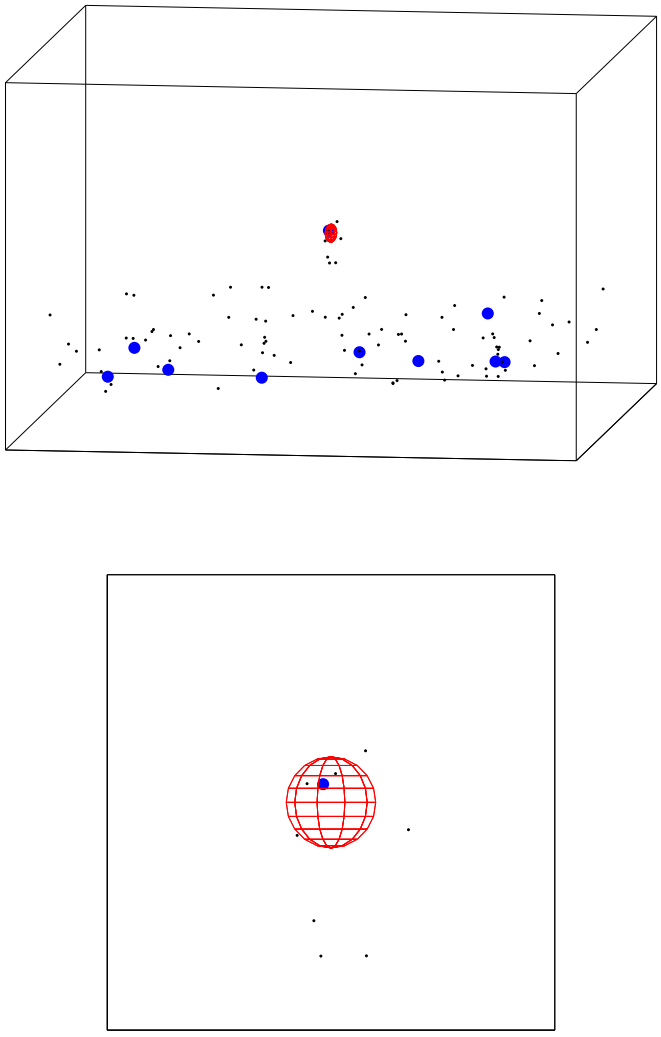

Figure 5: Scenario 1c.

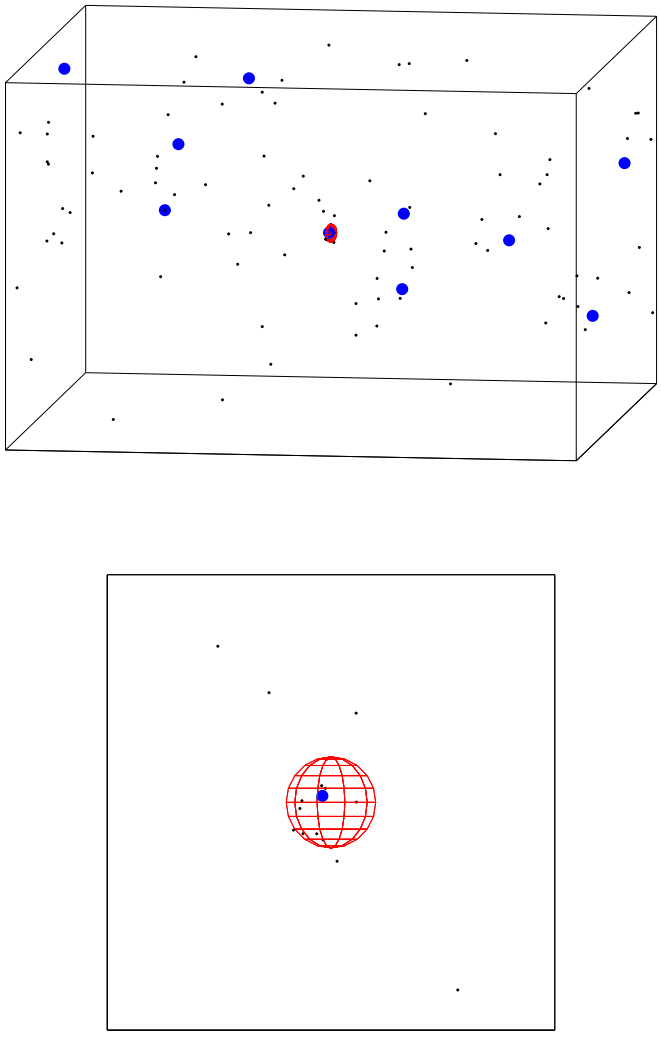

Figure 6: Scenario 2a.

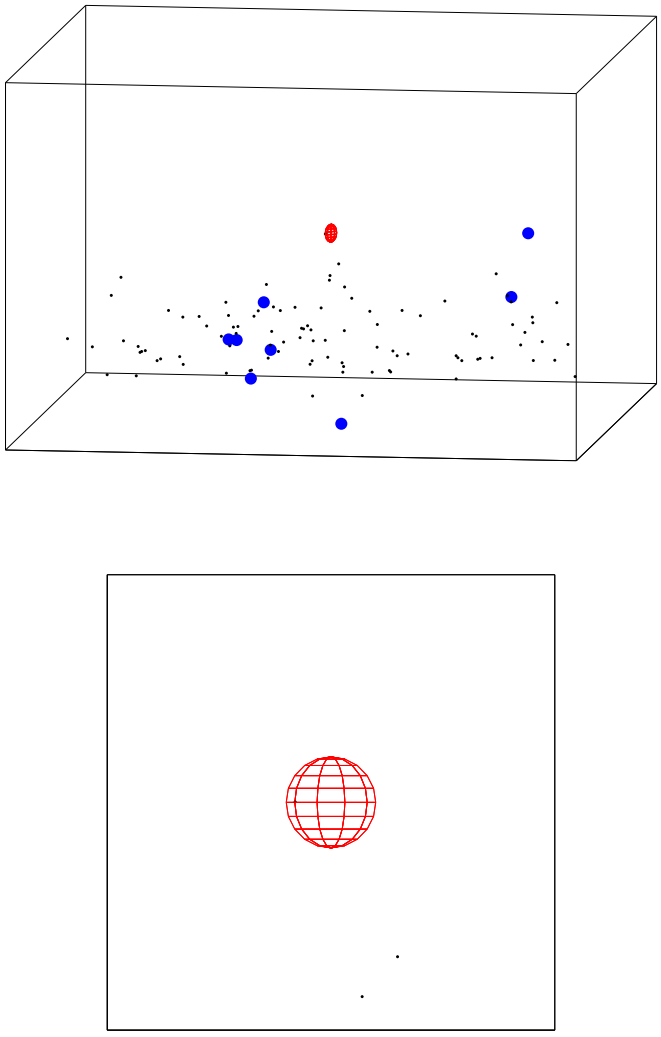

Figure 7: Scenario 2c.

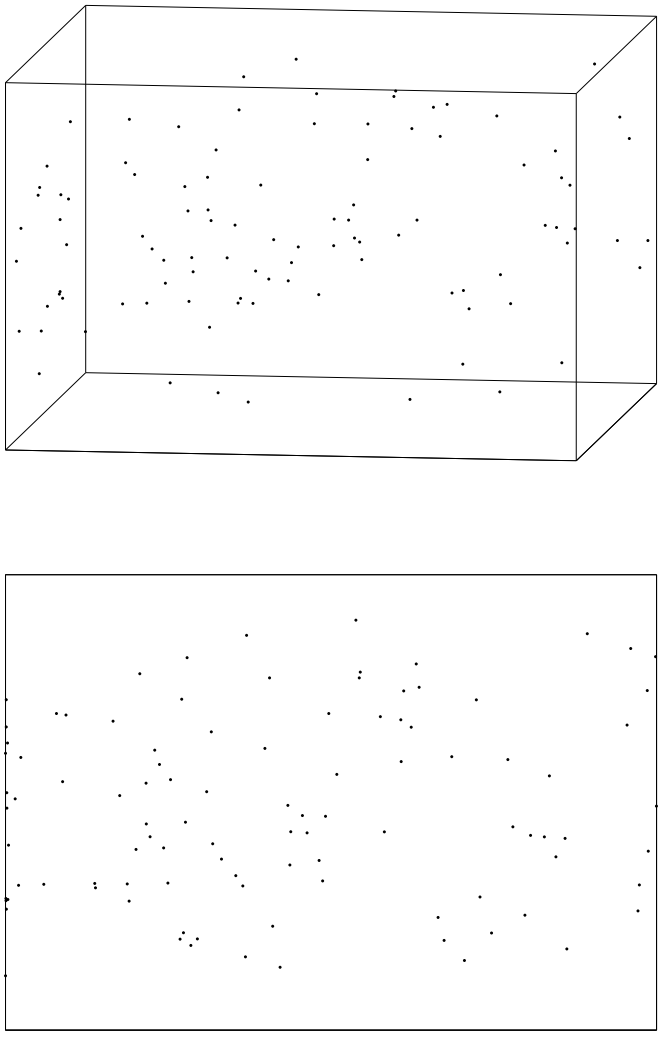

Figure 8: Scenario 6a.

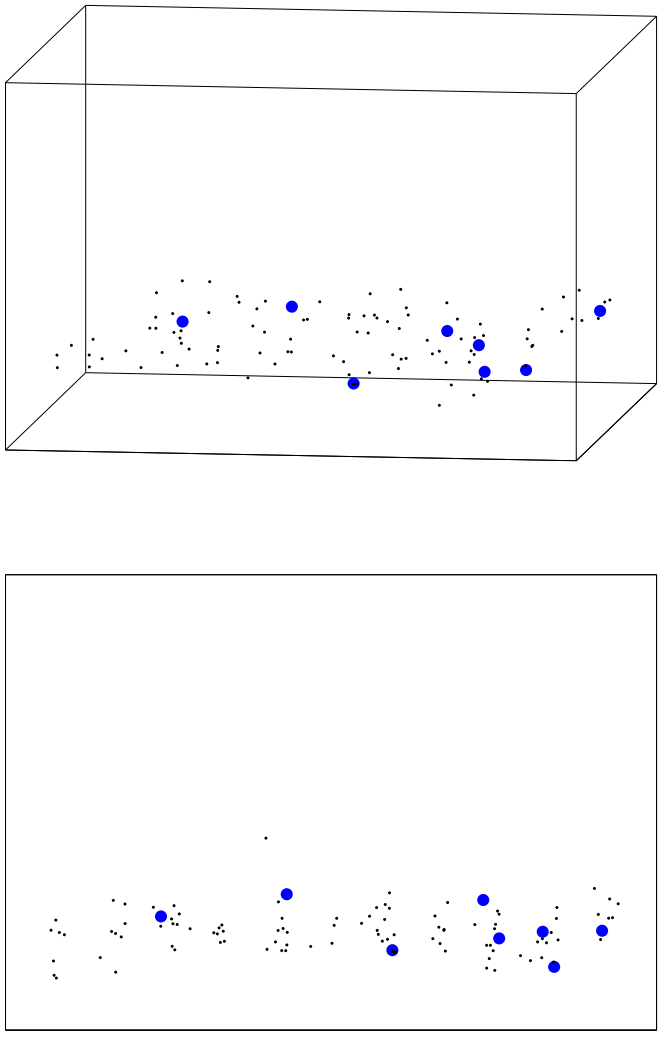

Figure 9: Scenario 6c.

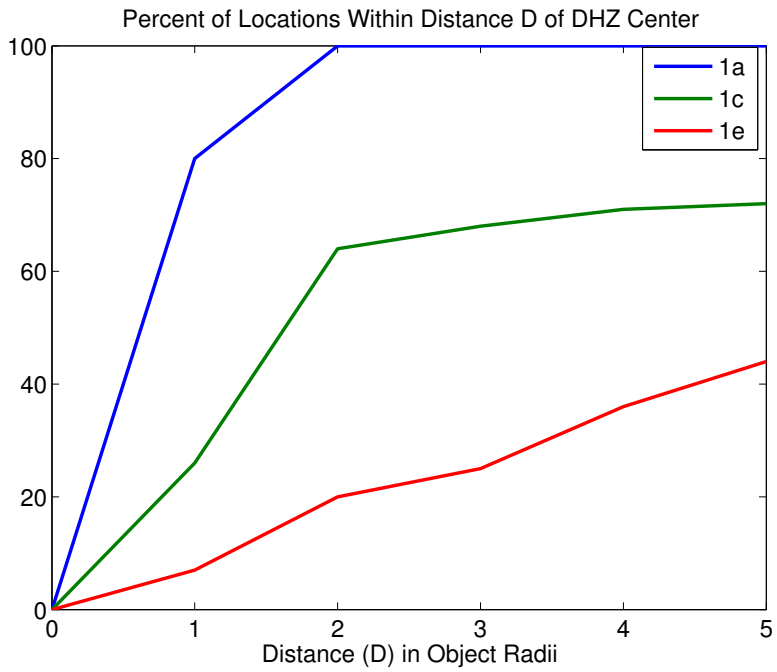

Figure 10: Clustering location summary for scenarios 1a, 1c, and 1e. The percent of best locations falling within a given distance of the DHZ location are plotted for a one minute experiment.

5 Likelihood Function Approach

Instead of relying on parameter selection in clustering techniques, suppose we develop a probability distribution associated with each muon of interest that tells us the likelihood that the muon underwent scattering at a particular container location (in the single-scatter approximation). Then the maxima of a weighted sum of these functions could reveal likely DHZ locations.

One sucn function assumes a Gaussian likelihood cloud based on PoCA locations and the modified distance function (Eq. 14).

$$U(s) = \sum_{j=1}^{k} S_j exp\left(-X_{sj}^2\right) \tag{15}$$

where S_j is the muon scattering significance and the index j sums over the subset of k muons considered relevant. Once again, and for the reasons outlined in the previous sections, we consider a relatively small subset of the total muon count consisting of those of largest S. The goal is to seek the local maxima of $U(s)$ and use these locations, or some subset of them, as likely locations of DHZ as test locations in a pretrained SVM.

Finding all the local maxima of $U(s)$ seems at first a computationally intensive task. Several important observations reduce the complexity greatly. First, the desired spatial resolution is only equal to or somewhat less than the characteristic size of any suspected DHZ. Second, we know a priori that the total number of maxima h is less than or equal to k. Third, each local basin in $-U(s)$ contains at least one point from the set p_j. I state this as a reasonable conjecture; to my knowledge, no proof exists. Fourth, the derivative of $U(s)$ is straightforward:

$$\nabla U = \sum_{j=1}^{k} 2S_j X_{sj} exp\left(-X_{sj}^2\right) \nabla X_{sj}. \tag{16}$$

$$\nabla X_{sj} = \left\{1 + \left[p\left(d_{sj}/a\right)^p - 1\right] exp\left[-\left(d_{sj}/a\right)^p\right]\right\} \nabla d_{sj}. \tag{17}$$

$$\nabla d_{sj} = \frac{1}{d_{sj}} \sum_{i=1}^{3} \left[\left(\alpha_{ji}^2 \left(\vec{s} - \vec{p_j}\right) \cdot \vec{e}_{ji}\right) \left(\left(\vec{e}_{ji} \cdot \hat{x}\right) \hat{x} + \left(\vec{e}_{ji} \cdot \hat{y}\right) \hat{y} + \left(\vec{e}_{ji} \cdot \hat{z}\right) \hat{z}\right)\right]. \tag{18}$$

Taken together, the above observations suggest the method of solution: using each p_j as an initial point, apply gradient ascent to find a local maximum to within the desired spatial tolerance.

The above approach is computationally intensive. A simpler approach that yields very similar results is to use the simpler potential function

$$U(s) = \sum_{j=1}^{k} \frac{S_j}{max(a, d_{sj})} \tag{19}$$

which looks like a graviational potential. The use of this function is difficult to defend. I use it here because results are obtained quickly and results are similar to those obtained by more defensible methods. Eventually, I will revert back to the use of Eq. 15 with quick search methods. But for the present consider the following results.

The merit of any likely-location method is that it returns a location close to the actual location of DHZ. I performed a likely-location analysis based on Eq. 19 for 100 realizations of scenarios 1, 2 and 3 with backgrounds a, c, d, and e for experiment times from 20 to 300 seconds. Figures 11, 12, and 12 show results in terms of the percent of best locations that fell within a given distance of the actual DHZ center location. Distances are measured in terms of DHZ radii (6.33 cm for the '1' cases, 4.22 cm for the '2' cases, and 5.00 cm for the '3' cases). The results are promising for the '1' cases and disappointing for the other cases as expected.

It should be noted that for all of these cases, unshielded radioactive DHZ will be detectable by other means. Locations of shielded DHZ should be easier to detect by the methods outlined in this report. The test cases represent challenging scenarios.

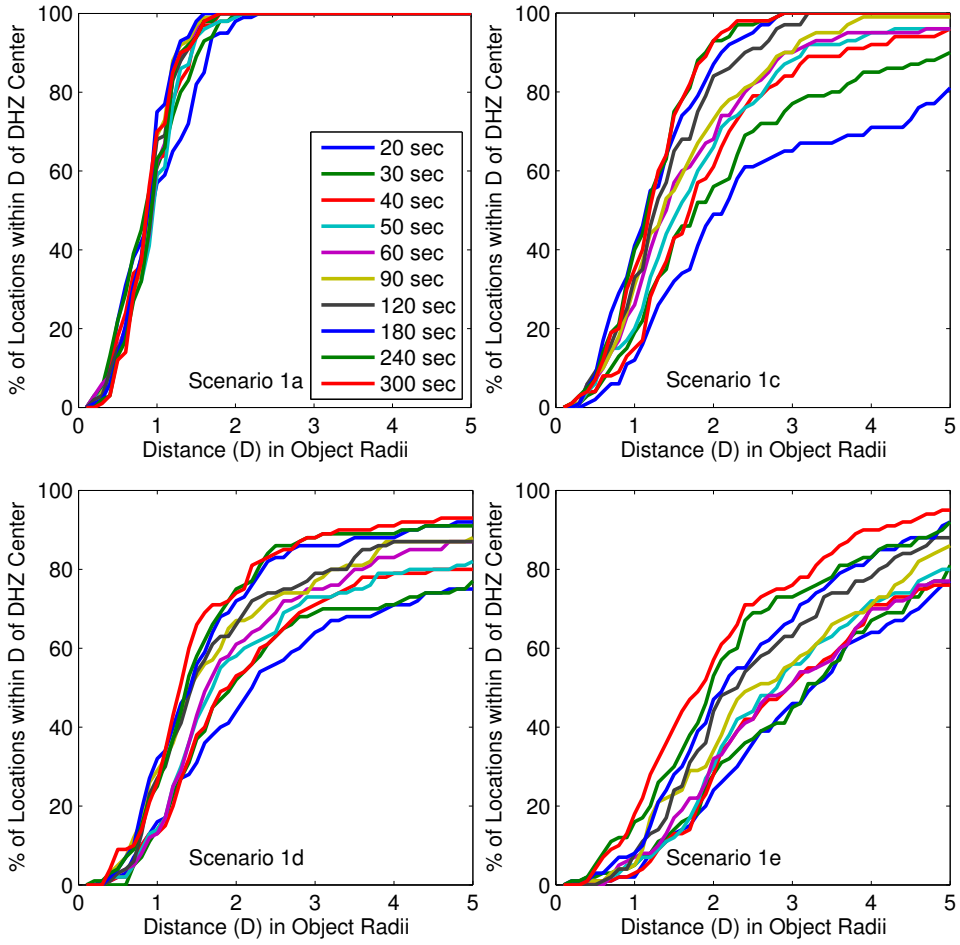

Figure 11: Likely location results for scenarios 1a, 1c, 1d, and 1e based on a test likelihood potential function. Best location results for different data collection times are shown as different colors.

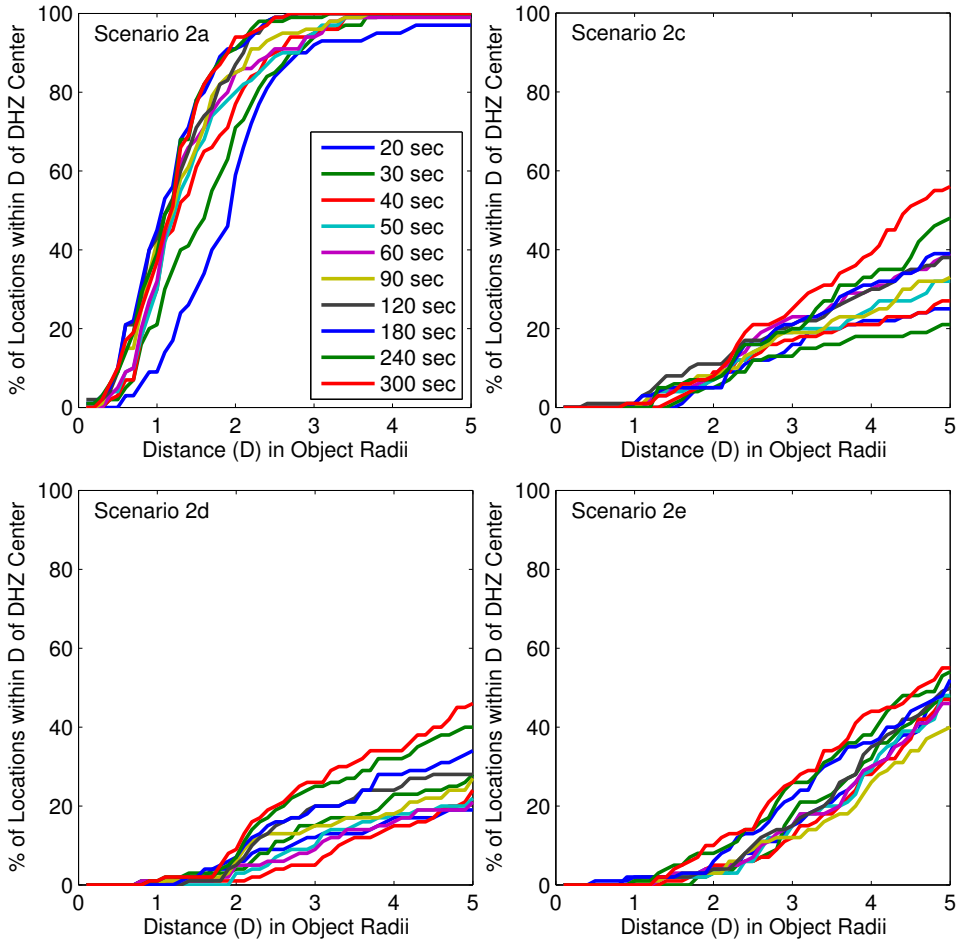

Figure 12: Likely location results for scenarios 2a, 2c, 2d, and 2e based on a test likelihood potential function. Best location results for different data collection times are shown as different colors.

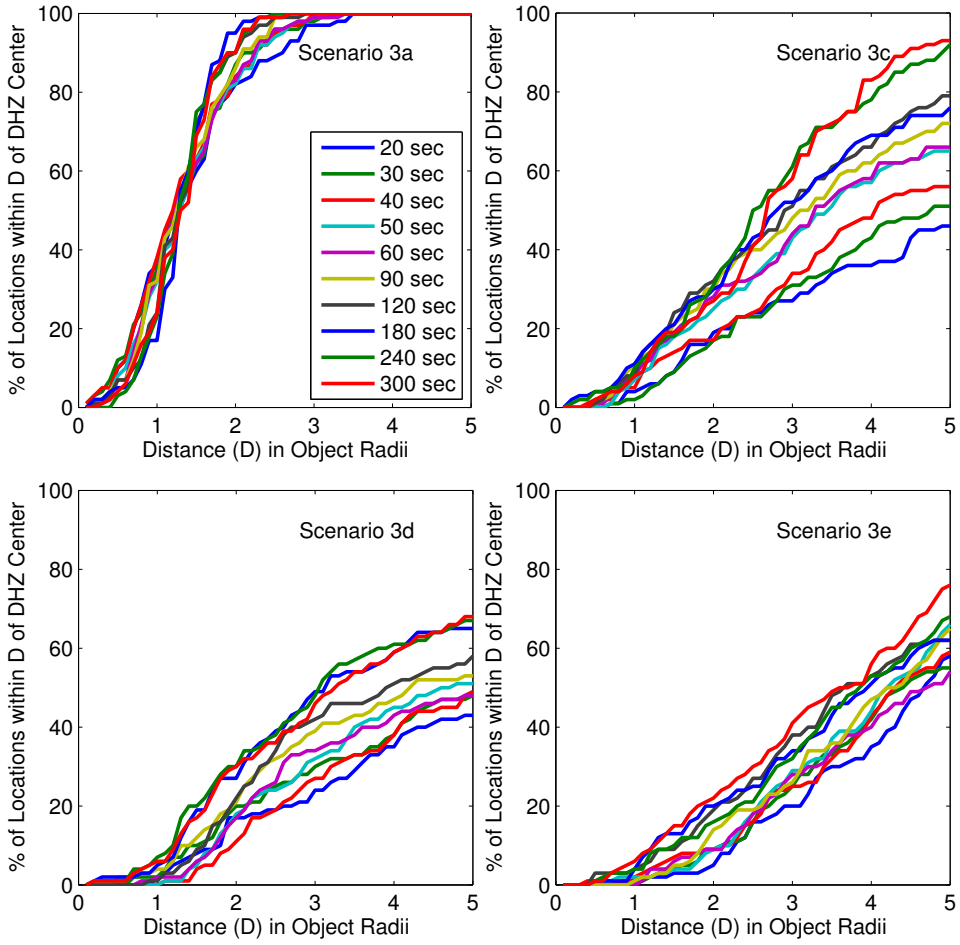

Figure 13: Likely location results for scenarios 3a, 3c, 3d, and 3e based on a test likelihood potential function. Best location results for different data collection times are shown as different colors.

6 Conclusion

This work is ongoing and should not be considered a final analysis. Areas of current work include (1) the inclusion of stopped muon information, (2) the effects of partial energy knowledge, (3) employing smart optimization schemes, and (4) building defensible likelihood functions.

This work was supported by the Los Alamos National Laboratory LDRD program and the United States Department of Energy.